DOCUMENT RESUME

ED 083 133 SP 006 889

AUTHOR Hartnett, Barbara M.; Rumery, Robert E.
TITLE Mark Chain Analysis of Classroom Interaction Data.
PUB DATE Feb 73
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, Louisiana, February 1973

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Interaction; *Interaction Process Analysis; *Preservice Education; *Student Reaction; *Student Teaching; Teacher Characteristics; Teaching Techniques

ABSTRACT

Classroom interaction data matrices obtained from a preservice training project in techniques of achieving divergent pupil responses were compared using a statistical procedure outlined by Darwin (1959). The procedure interprets interaction sequences as realizations of Markov chains. Contrary to results of an earlier study (Pena, 1972), interaction matrices were found to satisfy the dependency assumptions for Markov chains. Significant differences were found between interaction sequences for classes at different levels and between trained and untrained student teachers. Within-teacher comparisons were not significant. Pena's conclusion that the tests are too powerful is criticized on logical and methodological grounds. (Authors)

MARKOV CHAIN ANALYSIS OF CLASSROOM INTERACTION DATA

Barbara M. Hartnett
Lincoln College
Lincoln, Illinois 62656
Robert E. Rumery
Illinois State University

A paper presented to the Advance Copy
Session, Division D, of the annual meeting of
the American Educational Research Association
February 25 - March 1, 1973
New Orleans, Louisiana

MARKOV CHAIN ANALYSIS OF CLASSROOM INTERACTION DATA*

I. Introduction

In 1961, Ned Flanders described a technique for classifying and quantifying sequences of verbal behavior in the classroom. Verbal behavior is classified by a trained observer and then coded into a square interaction matrix. Each entry in an interaction matrix represents the frequency with which the row category is followed by the column category. Little work has been done in the quantitative analysis of classroom interaction data, aside from the derivation of several ad hoc indices such as the direct-indirect influence ratio. In 1959, Darwin derived a series of likelihood ratio criteria for comparing two or more realizations of Markov, or one-dependent probability chains. Darwin illustrated their use with some of Flanders' early classroom interaction data. In order for Darwin's tests to be applicable, however, the chains must possess the properties of a Markov chain.

Based on the likelihood ratio criteria developed by Hoel (1954) to test the length of dependence of a probability chain, Peña (1972) reported that the chains resulting from classroom interaction observations do not satisfy

*The authors wish to acknowledge the cooperation of Dr. Charles E. Gray and Dr. Richard C. Youngs for contributing the data analyzed in this paper. Data were originally collected for Experimental Project 178, "Preservice Teacher Training and Creativity in Problem Solving: A Developmental Investigation," August, 1971.

the requirements of a Markov chain; that is, Peña concluded that the probability of occurence of any verbal classroom behavior appears to depend, not on the immediately preceding behavior, but rather on the two preceding behaviors. Peña also concluded that Darwin's tests were too powerful because the tests often identified interaction matrices as different when the matrices resulted from observations of "educationally homogeneous" classrooms. Her criteria for identifying "educationally homogeneous" classrooms, however, were comparable mean scores on a math achievement test which are affected by many things other than verbal interactions within a classroom.

The results reported by Peña concerning the Markovian properties of interaction data are misleading. In order to obtain long chains for a single teacher, she combined data across five different subject areas; in order to obtain long chains for a single subject area, she combined data across five different teachers. Such additive procedures would be warranted only if differences in interaction matrices were shown to be independent of differences in subject matter and differences in teachers. Furthermore, by combining chains across teachers and across subject areas, Peña analyzed chains which ranged in length from 2,398 to 11,756 tallies (2-10 hours). Although the question of what constitutes an "observation" is somewhat ambiguous, it seems unlikely that continuous interaction sequences of this length would occur naturally in ordinary classroom settings.

II. Method

In the analysis reported here chains ranging in length from 167-544 tallies (8-27 minutes) were first tested for length of dependence, u, (Hoel, 1954), then tested for the equality of transition and occupation probabilities using the four likelihood ratio criteria derived by Darwin (1959). The data were obtained from a training project which had as its intent the establishment of teaching skills resulting in increased divergent pupil production in problem-solving.

Five sets of data were selected involving four student-teachers participating in the training project. Two of the teachers were observed in kindergarten classes, one a member of a training group, the other a control. Two of the teachers were observed in twelfth grade classes, again representing a training and a control group. A second observation for the control kindergarten teacher was included to provide within-subject comparisons. Observations were recorded on videotape and coded by a single observer who verified doubtful classifications by repeated viewings. The chains of behaviors resulting from these observations were each coded into a 9 x 9 matrix. Original observations were based on a more complex coding system (Amidon, Amidon and Rosenshine, 1969), but many categories had no entries in any row or column and were subsequently eliminated; other subcategories had few entries and were combined with closely related subcategories in order to simplify the analysis.

insert table 1

Hoel's test for the length of dependence (u) of a probability chains employs a likelihood ratio criterion distributed as chi-square. The Hoel procedure involves choosing a probable length of dependence, testing for significance and then decreasing u by one and testing for significance again. The procedure is stopped at the point where u is not significant but u-1 is significant. Each of the five interaction chains used in this analysis was tested for two-dependence and one-dependence. In all five cases the X^2 criteria (converted to Z) for two-dependence were not significant, while the X^2 criteria for one-dependence were significant.

insert table 2

These results indicated that each of the five chains used in this analysis is a realization of a Markov chain, and consequently, that the Darwin tests would be applicable.

The Darwin procedure provides four likelihood ratio criterion tests for comparing the equality of any number of realizations of a Markov chain. However, the present analysis considered only the case in which two realizations are tested for equality. Briefly, Darwin's four likelihood ratio criteria test the equality of:

1. two complete sets of transition probabilities, p_{jk}.
2. two off-diagonal sets of transition probabilities, p_{jk}, regardless of the diagonal values, p_{jj}.
3. two diagonal sets of transition probabilities, p_{jj}, regardless of the off-diagonal valnes, p_{jk}.
4. two sets of occupation probabilities, P_j, or the probability of occurence of a behavior in any category.

The criteria values of tests 2. and 3. are additively equal to the criterion value of test 1. Test 4, for the equality of two sets of occupation probabilities may seem redundant since the P_j values are related to the p_{jk} values. However, Darwin points out that it is possible for two realizations of a Markov chain to differ so slightly that the difference will not be detected by test 1, but that the particular functions of these differences as reflected in the P_j may result in the significance of the criterion value of test 4.

The four Darwin tests were applied to all six possible between-teacher comparisons and the one within-teacher comparison.

III. Results and Discussion

All four likelihood ratios were significant ($p > .001$) for the six between-teacher comparisons.

insert table 3

In these six comparisons, between 61-87% of the first criterion value was accounted for by test 2, indicating that these matrices had more pronounced differences in their off-diagonal entries than in their steady-states. In the within-teacher comparisons, the first three tests revealed no significant overall differences ($p < .10$), but the fourth likelihood ratio, testing the equality of occupation probabilities was significant ($p > .001$). These results suggest that interaction matrices may reflect within-subject consistency and that general activity patterns are essentially the same for a single teacher. Occupation probabilities, however, reflect differences in the time spent in each category of behavior, and such time differences may vary from situation to situation for a particular teacher.

Another noteworthy feature of the data is that the criteria values associated with between-grade, within-training comparisons were greater in magnitude than the criteria values associated with between-training, within-grade comparisons. Thus the data confirm the common sense assumption that grade level differences are more important than training group differences to the patterns of verbal behavior represented in an interaction matrix. In the original study from which the data were obtained, the grade level effect was also more pronounced than the training group effect in regards to the number of hypotheses generated per minute (hpm) during an observation. Rank order correlations (Kendall's Tau) were computed between each of the four sets of likelihood ratio criteria and the absolute difference in the number of hypotheses per minute (hpm) for each of the six between-teacher pairs.

insert table 4

Since the magnitude of a likelihood ratio criterion represents the degree of discrepancy between two realizations of a Markov chain, these correlations suggest that the greater the difference in the off-diagonal entries of two interaction matrices, the greater the difference in the number of hypotheses generated per minute in the two classrooms. A relatively strong relationship also appears to exist between differences in the total proportion of time spent in any behavioral category and differences in the hpms for two classrooms.

Peña also pointed out that one factor influencing the magnitude of a likelihood ratio criterion is the length of the chains used in an analysis. Rank order correlations between the total length of each pair of realizations and the four sets of likelihood ratio criteria were also computed and results tended to confirm Peña's statement, with the exception of likelihood ratio criterion 2, the test for the equality of off-diagonal probabilities.

insert table 5

Fewer entries in the off-diagonal cells appear to be somewhat related to large criterion values.

There is reason to believe that in sequences of ordinary lengths, chains derived from interaction analysis observations are one-dependent or Markovian in nature and consequently the procedure outlined by Darwin may be a useful method of quantitative analysis for the dynamics of classroom behavior. Differences in the magnitude of likelihood ratio criteria compared across grades and across treatment groups indicate that grade level has a greater influence on general patterns of verbal interactions than does training. Since within-subject comparisons were not significant in three of the four tests, a Markov analysis of classroom interaction data seems a potentially suitable method for describing the stable characteristics of a single teacher, and adding chains of observations across different teachers should not be done without previous testing to assure that all teachers' characteristics are similar.

Although the procedures used in this anlysis do offer a foundation for further theoretical work in the study of teaching, several issues of practical and theoretical interest remain unresolved. In this study, a relationship was found between differences in interaction sequences and the number of hpms, but analysis by a Markovian model was a post hoc procedure: data collection procedures were not designed to examine, or even reveal, relationships of this kind. Further theoretical attention should be given to appropriate choice of criteria. The question raised by Peña's conclusion that the Darwin tests are too powerful also remains unresolved and is confounded by the apparent relationship between the length of the chains used in the analysis and the resulting criterion value. An answer to this question would seem to depend upon a combination of a theory of teaching relatable to a Markov model and Monte Carlo studies of error rates for selected transition and occupation probability parameter values.

REFERENCES

Darwin, J. H. Note on the comparison of several realizations of a Markoff chain. Biometrika, 1959, 46, 412-419.

Ferguson, G. A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1966.

Flanders, N. A. Interaction analysis: a technique for quantifying teacher influence. Paper read at annual meeting of the American Educational Research Association. Chicago, 1961.

Gray, C. E. & R. C. Youngs. Instructional strategies for creative hypothesizing: a training program. Final report for Teacher Education Project 178, Illinois State University. Normal, Illinois.

Hoel, P. G. A test for Markoff chains. Biometrika, 1954, 44, 168-177.

Peña, D. The assumption of a Markovian chain model for interaction analysis. Paper read at annual meeting of American Educational Research Association. Chicago, 1972.

APPENDIX OF TABLES

TABLE 1

EXPANDED INTERACTION ANALYSIS CATEGORY SYSTEM

TEACHER TALK	1.	ACCEPTS STUDENT FEELINGS 1a - Acknowledges feelings 1c - Clarified feelings 1r - Refers to similar feelings of others
Combined	*2.	PRAISES 2w - Without criteria 2p - With public criteria 2p - With private criteria
	*3.	ACCEPTS STUDENT IDEAS 3a - Acknowledges ideas 3c - Clarifies ideas 3s - Summarizes ideas
	4.	ASKS QUESTIONS *4f - Factual questions 4c - Convergent questions *4d - Divergent questions *4e - Evaluative questions
Combined	*5.	LECTURES 5f - Factual lecture 5m - Motivational lecture 5o - Orientational lecture 5p - Personal opinion lecture
	6.	GIVES DIRECTIONS 6c - Cognitive directions 6m - Managerial directions
	7.	CRITICIZES 7w - Without criteria 7p - With public criteria 7p - With private criteria
STUDENT TALK	8.	STUDENT TALK, PREDICTABLE *8f - Factual student talk 8c - Convergent student talk
Combined	*9.	STUDENT TALK, UNPREDICTABLE 9d - Divergent student talk 9e - Evaluative student talk 9i - Student-initiated talk

*Used in present analysis

Table 1, Continued

NO TALK	*10.	SILENCE OF CONFUSION
Combined		10s - Silence 10c - Confusion 10 - A change of speakers in student-to-student interaction, and the beginning and end of a coding sequence in matrix construction.

From Edmund Amidon, Peggy Amidon, and Barak Rosenshine, SKIT Work Manual, Minneapolis: Association for Productive Teaching, 1969, p. 13.

TABLE 2

RESULTS OF THE HOEL TEST FOR THE LENGTH OF DEPENDENCE

OF A PROBABILITY CHAIN

TEACHER	GRADE	EXPERIMENTAL GROUP	LENGTH OF CHAIN (No. of tallies)	u	X^2	df	Z*
1	Kindergarten	Control	325 325	2 1	241.91 344.19	576 72	-11.93 14.28
1	Kindergarten	Control	167 167	2 1	146.73 228.32	576 72	- 7.63 9.62
2	Kindergarten	Training	307 307	2 1	191.89 293.96	576 72	-14.34 12.29
3	12th	Control	323 323	2 1	234.83 374.59	576 72	-12.26 15.41
4	12th	Training	544 544	2 1	93.57 816.08	576 72	-20.25 28.44

* $Z = \sqrt{2X^2} - \sqrt{2df-1}$. X^2 converted to Z due to df $>$ 70 .

TABLE 3

RESULTS OF DARWIN TESTS FOR THE EQUALITY OF

TWO REALIZATIONS OF A MARKOV CHAIN

	WITHIN-GRADE COMPARISONS		BETWEEN-GRADE COMPARISONS	
	Homogeneous Treatment	Heterogeneous Treatment	Homogeneous Treatment	Heterogeneous Treatment
LRC_2-off-diagonal	69.966 (KNT)	103.450 (KT vs. KNT)	150.328 (KT vs. 12T)	148.211 (12T vs. KNT)
	(within)	93.967 (12T vs. 12NT)	133.223 (KNT vs.12NT)	160.129 (KT vs. 12 NT)
LRC_3-diagonal	9.960 (KNT)	29.287 (KT vs. KNT)	82.000 (KT vs. 12T)	94.297 (12T vs. KNT)
	(within)	43.930 (12T vs. 12NT)	60.612 (KNT vs. 12NT)	23.622 (KT vs. 12 NT)
LRC_1-total	79.926 (KNT)	132.736 (KT vs. KNT)	232.327 (KT vs. 12T)	242.507 (12T vs. KNT)
	(within)	137.898 (12T vs. 12NT)	193.834 (KNT vs. 12NT)	183.751 (KT vs. 12NT)
LRC_4-occupation	30.615 (KNT)	47.501 (KT vs. KNT)	89.416 (KT vs. 12T)	94.416 (12T vs. KNT)
	(within)	50.945 (12T vs. 12NT)	37.323 (KNT vs. 12NT)	48.722 (KT vs. 12NT)

TABLE 4

RANK ORDER CORRELATIONS BETWEEN ABSOLUTE DIFFERENCES IN HPMS OF ALL BETWEEN-TEACHER COMPARISONS AND THEIR RESPECTIVE LIKELIHOOD RATIO CRITERIA

$hpm_b - hpm_a$

A. LRC

Criteria Values	1	2	3	4
1	-	1.17	.81	1.41
2	232	-	.36	.24
3	133	243	-	.60
4	184	138	194	-

$\tau = .07$

B. LRC_2

Criteria Values	1	2	3	4
1	-	1.17	.81	1.41
2	150	-	.36	.24
3	103	148	-	.60
4	160	94	133	-

$\tau = .60$

C. LRC_3

Criteria Values	1	2	3	4
1	-	1.17	.81	1.41
2	82	-	.36	.24
3	29	94	-	.60
4	24	44	61	-

$\tau = -.33$

D. LRC_4

Criteria Values	1	2	3	4
1	-	1.17	.81	1.41
2	89	-	.36	.24
3	48	94	-	.60
4	49	51	37	-

$\tau = .33$

TABLE 5

RANK ORDER CORRELATIONS BETWEEN TOTAL LENGTHS OF BETWEEN-TEACHER PAIRS OF CHAINS AND THEIR RESPECTIVE LIKELIHOOD RATIO CRITERIA

Total Length of Pairs

A. LRC_1

		1	2	3	4
Criteria Values	1	-	840	627	628
	2	232	-	868	876
	3	133	243	-	656
	4	184	138	194	-

$\tau = .48$

B. LRC_2

		1	2	3	4
Criteria Values	1	-	840	627	628
	2	150	-	868	876
	3	103	148	-	656
	4	160	94	133	-

$\tau = -.20$

C. LRC_3

		1	2	3	4
Criteria Values	1	-	840	627	628
	2	82	-	868	876
	3	29	94	-	656
	4	24	44	61	-

$\tau = .48$

D. LRC_4

		1	2	3	4
Criteria Values	1	-	840	627	628
	2	89	-	868	876
	3	48	94	-	656
	4	49	51	37	-

$\tau = .60$